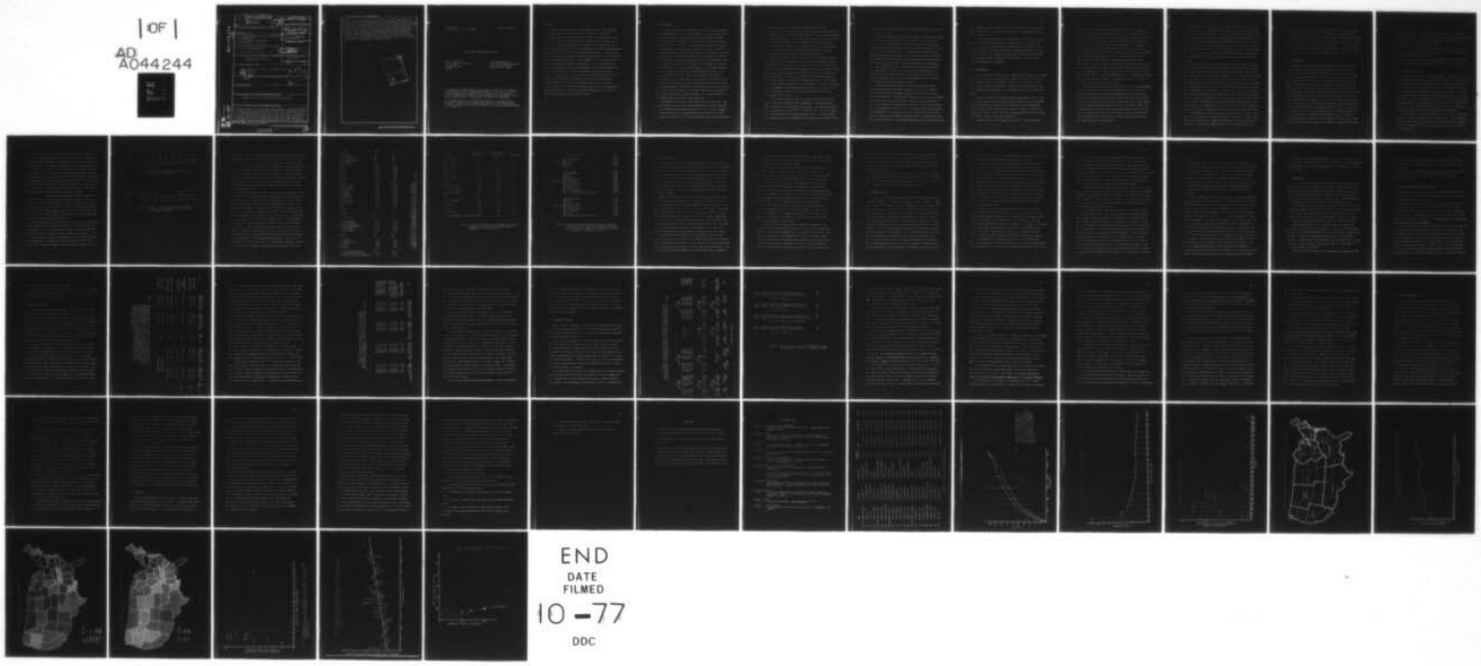
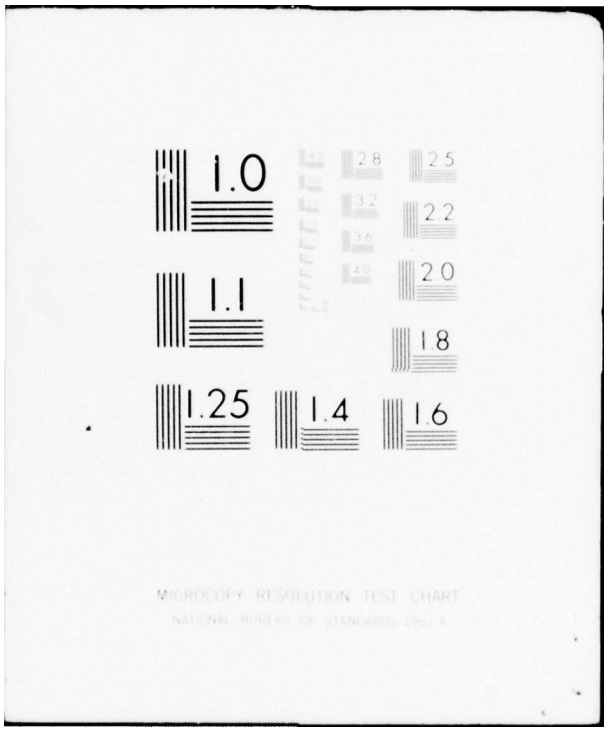
AD-A044 244 WEST VIRGINIA UNIV MORGANTOWN DEPT OF SOCIOLOGY AND --ETC F/G 5/11
THE REVERSE SMALL-WORLD EXPERIMENT.(U)
AUG 77 P D KILLWORTH, H R BERNARD N00014-75-C-0441
UNCLASSIFIED KB-114-77 NL



END
DATE FILMED
10-77
DDC


1.0
1.1
1.25
1.4
1.6
2.8
2.5
3.2
2.2
3.6
2.0
4.0
1.8
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963-A

AD A044244

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: (14) KB-114-77
2. GOVT ACCESSION NO.: (12) 10
3. RECIPIENT'S CATALOG NUMBER:
4. TITLE (and Subtitle): (6) The Reverse Small-World Experiment.
5. TYPE OF REPORT & PERIOD COVERED: (9) Interim rept.
6. PERFORMING ORG. REPORT NUMBER:
7. AUTHOR(s): (10) Peter D. Killworth (University of Cambridge); H. Russell Bernard (West Virginia University)
8. CONTRACT OR GRANT NUMBER(s): (15) N00014-75-C-0441-P00001
9. PERFORMING ORGANIZATION NAME AND ADDRESS: H. Russell Bernard, Dept. of Sociology and Anthropology, West Virginia University, Morgantown, WV 26506
10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS:
11. CONTROLLING OFFICE NAME AND ADDRESS: ONR-Code 452, Arlington, Virginia 22217
12. REPORT DATE: Aug[illegible] 1977
13. NUMBER OF PAGES:
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): West Virginia University, Morgantown, WV 26506
15. SECURITY CLASS. (of this report): Unclassified
15a. DECLASSIFICATION/DOWNGRADING SCHEDULE:
16. DISTRIBUTION STATEMENT (of this Report): Approved for public release, distribution unlimited
(12) 59 p.
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report):
18. SUPPLEMENTARY NOTES:
19. KEY WORDS (Continue on reverse side if necessary and identify by block number): networks, sociometrics, social structure, small-world
20. ABSTRACT (Continue on reverse side if necessary and identify by block number):

This paper considers the reverse small-world technique. Many "starters" are presented with a very long list (1267) of "targets." Starters are instructed in the small-world experiment and asked to write down their choice for the first link in the chain from them to each of the 1267 targets. The location, occupation, sex, and ethnic background of targets was given, and starters indicated which (if any) of these pieces of information made them think of their choice. Seven conclusions are drawn from the data: (1) A mean of 210

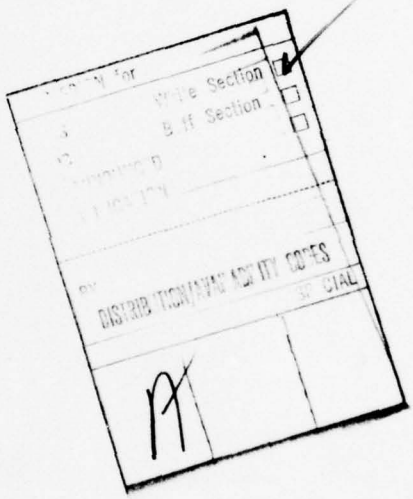
(Cont on p 1473 B)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601 |

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

(cont fr p 1473A)

choices control the world, but this number is an underestimate; only 35 choices control half the world, however. (2) There appear to be three distinct and mainly nonoverlapping, networks controlling U.S., local, and foreign areas; (3) Choices are mainly friends and acquaintances, with strong cleavages by sex; (4) Location is the usual reason for choices, with occupation second most popular; (5) The decision as to which choice is made depends primarily on the occupation of the trigger, and secondly on the distance (near/far) from Morgantown, West Virginia, where the experiment took place; (6) Half the states in the U.S. are controlled by a single person when choices are made on the basis of location, and (7) As found in previous experiments, the accuracy of informants in estimating their networks is low.



ONR Code 452
Contract #N00014-75-C-0441-P00001

Report # KB-114-77

THE REVERSE SMALL-WORLD EXPERIMENT*

Peter D. Killworth
Dept. of Applied Mathematics
Silver Street
Cambridge CB3 9EW
England

H. Russell Bernard
Dept. of Sociology/Anthropology
West Virginia University
Morgantown, WV 26506

* This work was supported under Office of Naval Research Contract #N00014-75-C-0441-P00001, Code 452. The opinions expressed in this paper are those of the authors and do not necessarily reflect the position of the supporting agency. Reproduction in whole or in part is permitted for any purpose of the United States Government. Distribution of this article is unlimited.

We are deeply indebted to our graduate assistant Ms. Ellen Cheung for her assistance in managing the data collection; and to Ms. Regina Phillips for assistance in data collection and for suggesting and asking the questions which led to Section 11.

Abstract

This paper considers the reverse small-world technique. Many "starters" are presented with a very long list (1267) of "targets." Starters are instructed in the small-world experiment and asked to write down their choice for the first link in the chain from them to each of the 1267 targets. The location, occupation, sex, and ethnic background of targets was given, and starters indicated which (if any) of these pieces of information made them think of their choice. Seven conclusions are drawn from the data: 1) A mean of 210 choices control the world, but this number is an underestimate. Only 35 choices control half the world, however. 2) There appear to be three distinct, and mainly nonoverlapping, networks controlling U.S., local, and foreign areas. 3) Choices are mainly friends and acquaintances, with strong cleavages by sex. 4) Location is the usual reason for choices, with occupation second most popular. 5) The decision as to which choice is made depends primarily on the occupation of the trigger, and secondly on the distance (near/far) from Morgantown, West Virginia, where the experiment took place. 6) Half the states in the U.S. are controlled by a single person when choices are made on the basis of location. 7) As found in previous experiments, the accuracy of informants in estimating their networks is low.

1. Introduction

It is by now obvious that the problem of measuring social structure is an extremely difficult one. In the past decade, it seems to us that subtle and unusual techniques for data acquisition have produced more valuable information about an individual's place in a given social group than more traditional sociometric methods. The small-world technique, due to Milgram, is a case in point. Since Milgram's original article (1967), a number of researchers have duplicated, verified and adapted the experiment in a variety of environments. (See Travers and Milgram, 1969; Lundberg, 1975; Bochner, Buker and McLeod, 1976; Hunter and Shotland, 1974; Korte and Milgram, 1970).

The small-world problem can be thought of as an attempt to define the incoming network of a preselected target person. (In this paper, "network" is used to connote those people whom a person knows and can turn to for various purposes.) A group of "starters" are asked to send a folder to a target person. If the starter does not know the target personally, then he is asked to send the folder to someone whom he believes has the best chance of knowing the target. The folder is passed along with the same instructions. Thus, a chain from each starter to the target is constructed. The group of people who eventually give or send the folder to the target constitute, in some undefined way, the target's "incoming network."

Two things stand out in the research done in this field so far: 1) the number of links in the chains is surprisingly small (on average, there are 5.25 intermediaries between any two people in the U.S.); and 2) the number of people who constitute the incoming network is also small (in the 1967 experiment, for example, of the 62 completed chains, only 26 final links were

required, and 3 final links accounted for nearly half the chains.)

There are two good reasons not to consider the small incoming network as the target's total (i.e., nondirectional) network. First, only 62 realizations produced the 26 final links. We can expect this number to rise as the number of chains rises; for obvious reasons, however, it should rise at an ever decreasing rate (with respect to the number of chains) and will, presumably, asymptote to a finite value. The asymptotic value would be the target's total incoming network; therefore, 26 must be an underestimate.

Second, 26 links are too few to function as a set of paths away from a starter in a small-world experiment. We can see this very simply: if each individual in the U.S. (population = 2×10^8) had 26 links to the rest of the world, his links' links would number at most (i.e., with no overlap) 26^2; their links (same restriction applying) would be 26^3; and so on. This means that at a distance of 6.2 links from an individual (5.25 intermediaries yields 6.25 links), a person could contact (via appropriate chains) 7×10^8 people. This number is clearly an overestimate on two grounds: 1) whatever "structure" is, it is composed of overlap between people's networks; an overlap of only 10% (so that powers rise by 0.9 at each remove) reduces the size of the contactable world to 9×10^7.

2) In the above calculation we have assumed that each person, at all stages of the chain, chooses the best intermediary to continue the chain. Suppose that only one "mistake" per chain occurred. This reduces the size of the contactable universe from $26^{6.25} = 7 \times 10^8$ to $26^{5.25} = 2.7 \times 10^7$. Thus, even a very small amount of error has a drastic effect. In fact, allowing for one error in a mean chain length of 6.25, one would need 38 links to reach

the entire United States; with two errors, this number becomes 90; and with 3 errors it reaches 360.

Of course, Milgram's experiment was not designed to exhaust the target's incoming network, but rather to find out how many links there are in the chains. But, to understand social structure, we do need to know how many links people have to the rest of the universe. There are three ways to get at this. The most obvious would be to repeat Milgram's experiment using hundreds (or thousands) of starters. This costly and logistically horrifying modification would still yield the incoming network of only one person -- the target. A way to study many individuals' networks is to ask people to keep track of whom they contact. Gurevich (1961) found that, on average, people come into contact with 500 other persons in a 100-day period. Of course, this does not mean that many of these contacts would be of use to people if they were starters in a small-world experiment, but it does give an indication of the sorts of numbers involved.

We wanted to design an experiment which would, hopefully, tell us exhaustively about many individuals' networks. To do this, we combined Milgram's and Gurevich's techniques to produce what we term the "reverse small-world" experiment. Instead of many starters and one target, we presented each starter with a long list of targets. For each target, we provided a variety of information, and we asked the starter to tell us to whom he would send the (now mythical) folder if he were initiating a small-world experiment. This experiment has the following trade-offs: all information about chain lengths is lost; and the experiment deals with people's cognitive rather than behavioral networks (i.e., to whom people <u>say</u> folders would be sent, as distinct

from whom they actually send folders to). On the other hand, in a very short time a vast amount of highly detailed information can be accumulated from a wide variety of starters.

This experiment also provided a convenient way of testing whether or not people understand the size and the characteristics of their networks. Our previous studies (Killworth and Bernard, 1976; Bernard and Killworth, 1977) have shown that, in very limited environments, people are inaccurate in their reports of their actual communication with others. We wanted to know whether this inaccuracy extends to people's global networks. Hence, the reverse small-world experiment was conducted.

2. The Experiment

First, a list of targets was needed. Since we were not going to ask people to actually send folders to their network, we prefer to think of the targets as "triggers." That is, they trigger a response in the form of a "choice" (i.e., one of the starters' network). Since "starters" really weren't starting anything, the anthropological term "informant" will be used throughout the paper.

The list of triggers had to satisfy several requirements. First, it had to be long; but how long? We wanted to provide informants with a fair number of triggers from various walks of life, locations, and ethnic backgrounds; and we wanted to ensure that there were sufficient triggers to exhaust each informant's network. Gurevich's (1961) data suggested that at least 500 would be required. Doubling this figure seemed safe.

Eventually a list of 1267 triggers was created. The first 1000 were

generated to fill a matrix based on sex (male and female), race (black and white), occupation (professional, white collar, craftsman, housewife), and location (big city, small town). This last variable was subjectively operationalized; a location was classed as "big" if, in our judgement, most Americans would recognize it by name. Seattle, Tallahasee, Houston, are examples. The scale developed by Duncan and Reiss (1961) was used to select occupations. Since a very large part of the world is made up of housewives, that occupation was assigned to 25% of the female triggers. In retrospect, this was not necessary, and a greater array of occupations should be used in future experiments. In general, larger states (Illinois, California, New York) are heavily represented (44, 57, and 44 occurrences), while smaller states, such as Alaska, Delaware, and Vermont occur less frequently (7, 10 and 15 triggers respectively). All 50 states, plus Washington, D.C. (4 triggers) were included in the final instrument. Somewhere in the gargantuan shuffle to create the instrument, a trigger got lost; this left 999.

In order to test whether ethnicity (rather than "race") influenced people's choices of a first link in a small-world experiment, a second list of 168 names was generated. These were clearly identifiable to most Americans as Slavic (i.e., Olga Zdrojewski); Spanish (Francisco Gonzalez): Italian (Maria Vaglienti); and Oriental (Wong Fuk Lam). The 144 Slavic names, Spanish names, and Italian names were equally divided by sex, and by "big" or "small" town. They were further divided into professionals and unskilled workers. The 24 Orientals were not divided by sex; it was assumed that most Americans could not tell the sex of, say, Wong Fuk Lam. For most purposes, black triggers are treated as ethnic in what follows.

Finally, a list of 100 triggers was created to represent "the rest of the world." These are exotic names from exotic places. Only small places were used, so that informants would generally have heard of only the trigger's country. The occupations of these last 100 triggers were evenly divided into professionals (i.e., chemical engineer) and lower class workers (i.e., mail clerk). The first 1167 triggers were presented to informants in essentially random order. The 100 "exotic" triggers were left to the end. Presumably, by this time (usually several days into the test-taking procedure) informants had decided who to use for all conceivable triggers in the world. As we will show below, this conjecture was premature.

A copy of the list (a specimen page of which is given in Figure 1) was presented to each of 58 informants who were chosen to cover a wide range of social and economic backgrounds. All lived in or near Morgantown, W.Va. Each informant was first instructed in the small-world experiment. For each trigger the informant provided the name of a choice (who was known well enough to be used in a small-world experiment), the choice's relationship (friend, mother, etc.), and a reason for making the particular choice. Four reasons were provided: location, occupation, ethnicity or race, and "other." In other words, if the reason given were location, then something about the location of the trigger and/or some knowledge about the location or life experience of the choice was involved in the selection of that choice. Informants were allowed to check more than one reason if they felt the need to. Overwhelmingly, informants chose only one reason. Where more than two reasons were checked ($<.01\%$ of choices), only two reasons were coded.

This provided us with two sets of data. We had a list of triggers, about which we knew occupation, sex, ethnicity, whether they lived in a large or small town, and location by state. Foreign countries were divided for analysis into five blocs: South America, Western Europe, Far East, Middle East, and

Eastern Europe. We also had, for each informant, a list of choices (one per trigger) about whom we knew a) sex, b) relationship to informant, and c) a reason for the informant making that choice. In addition, we also had some background information on the informants (age, sex, income, religion, etc.), together with their responses to three questions which we asked 1 - 4 months after the test was over: "How many different choices did you make on the test?"; "Which choice did you use the most?"; "Which three choices did you use the most?"

3. The Informants

Fifty-eight persons completed the questionnaire in 8 hours, on average. (Informants came to our office to take the test, but were under no time pressure. They returned as often as required to finish.) Another 12 found the task too tedious and dropped out. The mean age was 36, s.d. = 15. There were 34 women and 24 men in the group; 29 were married, 4 were divorced, 20 were never married, and 5 were either widowed or separated. Eighteen came from big cities; 16 from small cities; 15 from small towns; and 5 were from rural backgrounds. (The city of origin of 4 informants is unknown.) Seventeen were employed full time; 20 part time; and 21 were unemployed. (All the unemployed were housewives, students, or retired.) Seventeen of the women were full-time housewives, and 7 were part-time housewives. The educational level of informants ranged widely. One completed only grade school; 6 high school; 8 completed high school and had vocational training; 14 completed at least some college years; 16 completed an undergraduate program; 12 completed a graduate program; and one had a doctorate. Seven had a yearly income of less than $2000. (These were all students.) Fifteen had yearly incomes between

$2000 - $5000; 5 from $5000 - $8000; 9 from $8000 - $11000; 3 from $11000 - $14000; 4 from $14000 - $17000; and 14 had family incomes beyond $17000 per year. Thirty were Protestant; 8 Catholic; 3 Jewish; 7 "other religion" (mostly foreign students); and 10 professed no religion.

Informants also provided an estimate of their "ethnicity." They created 13 categories; 40 considered themselves "American;" 18 said they were Spanish American, Black, Italian American, etc.

Personal data were stored for each informant in the chunks as indicated here. For example, education was coded 1-7, from grade school to doctorate.

4. The Number of Choices

Of immediate interest is the size of a person's total network. Figure 2 shows a cumulative histogram of the mean number of different choices accumulated as the number of triggers increased towards 1267. The change in gradient at 1167 is obvious. It results from the introduction at this point of the 100 foreign triggers. Two observations can be made about this curve. First, the error bars are enormous. Second. the curve has not asymptoted to a constant value before the foreign triggers were introduced. Unpleasantly, this implies that 1167 triggers are insufficient to produce the entire U.S. network of an informant. A rough estimate suggests that at least 2000 U.S. triggers, and perhaps 500 foreign triggers, are necessary to exhaust the network.

The mean number of different choices was 210, s.d. = 168. The range was 43 to 1131. Crude extrapolations of the curve in Figure 2 would suggest an asymptotic value of about 250, plus 20 - 25 to account for foreign triggers.[1] (Some foreign triggers will be accounted for by those who also account for persons in the U.S.)

The shape of the curve in Figure 2 is also of interest. Since the triggers are arranged randomly, the shape of the curve will be the same, no matter what orders the triggers are presented in. As a test, the dotted line shows the equivalent for triggers in the reverse order. It lies well within the error bars of the original curve, except for the first hundred triggers (now foreign), where its rate of growth is significantly slower. (In fact, an informant's choices which control the foreign world are highly significantly (0.1%) fewer and more general than those used for the U.S. - an average of 34.9 choices handle the 100 foreign triggers, compared with 55.8 for the first 100 U.S. triggers.) This apart, the strong similarity of the curves suggests that its shape is universal, at least for the sample studied.

One can also compute how many choices are required to account for a given fraction of the world (remembering with caution the lack of asymptote). This is shown in Table 1. Note that remarkably few choices account for a great deal of the world: in particular, 34 choices account for 50%, and 125 choices (half the average total choices) account for 90% of the world. If only the foreign triggers are considered, 8.5 choices account for 50% and 26 for 90% of the world beyond the U.S.

Is there any structure in the curve in Fig. 2? This was tested by a model similar to that of Pool and Kochen, referred to in Travers and Milgram (1969). The assumption is that each individual has N friends, each of whom accounts for $1/N^{th}$ of the world. A friend is chosen randomly for any trigger. It is then a trivial matter to produce the expected histogram of cumulative friends, to compare with Fig. 2. However, even when tuned optimally, the fit was extremely poor. Attempts to improve this by subdividing "friends" into sub-

Percentage	10	20	30	40	50	60	70	80	90
Mean	3.2	8.5	15.5	24.0	34.2	47.2	64.0	87.0	125.4
S.D.	3.5	16.4	33.4	50.8	68.4	88.1	108.9	130.0	151.9

TABLE 1a. Number of choices required to handle a given percentage of the world.

Percentage	10	20	30	40	50	60	70	80	90
Mean	1.57	2.64	4.34	6.07	8.43	11.48	15.28	19.88	25.86
S.D.	0.92	1.91	3.59	5.47	7.59	10.02	12.74	15.99	19.22

TABLE 1b. Number of choices required to handle a given percentage of the world (limiting consideration to the foreign triggers.

groups (e.g., "good" and "poor") with different proportions of network size did not improve this. We must conclude, then, that the curve in Fig. 2 is not random and does represent a facet of the informants' mean social structure.

Some choices are much more efficient at controlling the world than others. Fig. 3 shows the percentage of the world controlled, on average, by the top 20 choices. The top choice stands out dramatically at just under 10% (the apparent discrepancy between this and Table 1 is due to the different averaging procedures used), and will be examined in detail in Section 7. Limiting consideration to foreign triggers, an informant's (usually different) top choice accounts for 21%.

The variation in the toal number of choices between informants is shown in Figure 4. It is noteworthy that the two very large numbers of choices were produced by an 69 year old woman, and a 19 year old student from Thailand. Such large variation suggests that there may be systematic features in the informants' backgrounds. Comparison of SES variables on informants with the number of choices they made yielded the following: only age, education and income had <u>any</u> effect on the number of choices. The effect of education was limited to males, and is both weak and nonlinear. Age and income together account for 16% of the variance in the number of choices (with a multiple correlation of 0.39^{*2}, and 2-way analysis of variance showing a similar result). Single correlations were choice-age 0.29^{*}, choice-income -0.18. In other words, older informants have more choices, except those in the higher income brackets, and the Thai student mentioned above. Since 16% of the variance is not a lot, there seems to be no overriding reason which accounts for the variation in number of choices, except individual life histories.

TABLE 2. Entries are the number of different choices used in the given category. The notation P ( ) indicates the number divided by the total number of choices made. (Not all fractions sum to unity, due to missing data.)

Variable	Mean	S.D.
male friends	69	85
female friends	42	38
male acquaintances	46	36
female acquaintances	19	16
male family	12	7
female family	11	7
mother	0.59	0.53
father	0.64	0.64
spouse	0.52	0.50
brother	1.14	1.21
sister	1.05	1.18
son	0.31	0.65
daughter	0.28	0.72
cousin	7.3	8.2
uncle	2.9	3.1
aunt	2.8	2.6
mother-in-law	0.38	0.52
father-in-law	0.33	0.51
daughter-in-law	0.12	0.38
son-in-law	0.14	0.44
grandmother	0.26	0.58
grandfather	0.29	0.73
granddauthter	0.09	0.47
grandson	0.05	0.29
sister-in-law	0.95	1.41
brother-in-law	1.21	1.63
niece	0.02	0.13
nephew	0.55	1.39
family	23	13
friends	117	137
acquaintances	70	54
P (male friends)	0.30	0.14
P (female friends)	0.20	0.09
P (male acquaintances)	0.23	0.13
P (female acquaintances)	0.10	0.06
P (male family)	0.08	0.06
P (female family)	0.07	0.05
males	127	104
females	72	48
P (friends)	0.50	0.18
P (acquaintances)	0.32	0.17
P (family)	0.14	0.10
P (male)	0.60	0.10
P (female)	0.36	0.10
friends and acquaintances	176	143
P (friends and acquaintances)	0.82	0.10
friends/(friends and acquaintances)	0.60	0.22

Category	Number used by males	Number used by females	Significance
male friends	95	50	*
male acquaintances	65	33	**
female family	8	12	*
acquaintances	90	56	*
sisters-in-law	0.3	1.4	**
brothers-in-law	0.7	1.6	*
nephews	0.13	0.85	*
P (female friends)	0.14	0.24	**
P (male acquaintances)	0.29	0.18	**
P (female family)	0.04	0.08	**
males	172	96	**
P (acquaintances)	0.38	0.29	*
P (family)	0.10	0.17	*
P (males)	0.69	0.54	**
P (females)	0.27	0.43	**
P (friends/(friends & acquaintances)	0.86	0.80	*

TABLE 3. Significant differences in the numbers of different categories of choices between male and female informants. P ( ) as in Table 2.

Correlation between		Value
age and	male family	0.28*
	female family	0.27*
	family	0.30*
	cousins	0.26*
	nephews	0.72**
education and	male family	-0.33**
	female family	-0.35**
	sons	-0.24*
	daughters	-0.26*
	daughters-in-law	-0.42**
	sons-in-law	-0.23*
	sisters-in-law	-0.35**
	brothers-in-law	-0.30*
	nephews	-0.42**
	P (male family)	-0.35**
	P (female family)	-0.38**
	P (friends/(friends+acquaintances))	0.42**
income and	female friends	-0.23*
	spouses	0.46**
	sons	0.28*
	fathers-in-law	0.22*
	sisters-in-law	0.28*
	brothers-in-law	0.45**
	P (male family)	0.31**
	P (family)	0.27*
	females	-0.24*
	P (friends/(friends+acquaintances)	-0.29*

TABLE 4. Significant correlations between numbers of different categories of choices and characteristics of informants. P ( ) as in Table 2. Education and income are blocked into 7 categories as defined in Section 3.

5. Types of Choices

Informants made essentially three different kinds of choices: friends, acquaintances, and family members (this latter being divided into 22 categories -- see Section 2). Table 2 gives the mean numbers of choices in each category. Friends and acquaintances account overwhelmingly for most of the choices (82%). This is in agreement with Travers and Milgram (1969) who found a figure of 86%. Male choices (127) are used much more than female choices (72) by both men and women. Nieces are hardly ever used; nephews, by contrast, are used by half the informants. Cousins are used quite frequently (over 7 different cousins per informant).

Tables 3 and 4 examine how the usage of different categories of choices varies with some characteristics of the informants. The most obvious difference between informants is their sex (Table 3 lists all significant cases). Males choose more males than females do (172 to 96)**. However, neither males nor females choose more females. Males are 2.5 times more likely to choose males than they are to choose females (0.69 to 0.27). However, females are not more likely to choose females than to choose males (0.43 to 0.54). This confirms, qualitatively, a result from Travers and Milgram (1969). They found that men were ten times as likely to send a document to other men than to women; whereas women were equally likely to send documents to men or women.

Males have more acquaintances (90) than do females (56), but neither males nor females have more friends. Males choose more male friends and more male acquaintances than do females. The data show that at least selected parts of the family would be used as first links by females more than by males. This is in line with the common wisdom that females are more likely to depend on

their family than are males, for contact with the world. This shows up quite explicitly: 17% of choices made by females were family members, whereas only 10% of the choices made by males were family.

Table 4 shows the dependence of choices on other characteristics of informants. There is a strong tendency to choose more family members as the age of the informant increases. (Obvious correlations, such as the tendency not to choose one's mother as one gets older, are omitted.) Furthermore, there is an even stronger tendency to choose fewer family as education of informant increases. Note here the strong positive correlation between education and the fraction of the informant's friends and acquaintances which are termed "friends" by an informant. (Let us call this the "friends fraction.") In other words, more highly educated persons tend to say that more of their nonfamily choices are their "friends" than do less educated persons.

There is a similar tendency for informants to choose family members (especially in-laws) more as the income of the informant increases. In addition the "friends fraction" decreases with income. (N.B. Because of the students in the sample, the traditional high correlation between education and income is not present in the data.)

Counterintuitively, those who identify themselves as "American," i.e., nonethnic, choose significantly* more family (25) than do ethnic groups (18). The same is true for female family but not, curiously, for male family. Since 6 of those informants who were coded among the "ethnic" group were foreign students, the data were re-run without them. We reasoned that since they were not in their own country, they would be less likely to use their family as choices. This, we assumed, would increase the use of family by those in the

"ethnic" category. Unfortunately, our assumptions were wrong, and removing the foreign students did not affect the curious finding above, although it did reduce its significance. It is also possible that families of informants with ethnic backgrounds may be seen (by informants) as having rather limited networks of their own.

Finally, the probability of choosing family (male family, female family, and total family) decreases monotonically with the population of the informant's natal home. In other words, persons from rural backgrounds are more likely to choose family than are persons from small towns; and so on.

6. Individual Networks

This section examines which "areas of the world" are handled by each choice made by an informant. For example, does one have "someone in Idaho?" -- namely, a choice which is perpetually selected for triggers in Idaho. Similarly, does one tend to rely on one's "dentist" or on a single Mexican American to reach all the triggers who are dentists or Mexican Americans? This will be examined in detail in Section 11; here we will describe, qualitatively, the features of the world which are handled by an informant's most frequently used choices. An example of the difficulty in being quantitative is easily given. Consider informant 11. His 7th most frequently used choice was almost always selected on the basis of location (55 times out of 56). On those occasions, this choice was used for triggers in Washington 20 times, Oregon 33 times, and Utah 2 times. Clearly, informant 11 had "his man in the Northwest;" but the choice was used overwhelmingly for just two states. The decision to remove Utah for descriptive purposes is purely subjective. Furthermore, all cases are not as

clearly defined as the one given. Despite these problems, we feel it is worthwhile to give what must be qualitative descriptions of these data.

Informants' top 8 choices were examined in detail. (Recall from Table 1 that the first 8 choices accounted for 20% of the world.) For 72% of the informants the top choice (quantitative details of which are given in Section 7) handled a recognizable set of states when chosen on the basis of location. When this occurred, the choice was used for a wide variety of locations. For descriptive purposes, the U.S. was divided into eight blocs as defined by the U.S. Office of Education (Fig. 5). In one case a choice covered all eight blocs, and for one informant, the top choice was selected for six of the eight blocs. Otherwise, chosen on the basis of location (typically two to three blocs), informants' top choices tended to be used for rather more restricted areas. Ten of the top choices were unambiguously selected on location for triggers from a single state.

Top choices are less frequently selected on the basis of occupation (only by 28% of the informants). This tendency is, in fact, true no matter which choice is examined (see Section 8). The most popular occupation for the top choice to handle is that of housewife (12% of informants). Four top choices handle other single occupations, one handled a wide range of occupations; one handled a pair of widely differing occupations; some handled only high-status occupations, others only low; and so on. No recognizable pattern emerges. Some top choices handle both locations and a specific occupation.

As one examines less frequently-used choices, a tendency toward more specialized functions for the choices emerges. Figure 6 shows the total number (over 58 informants) of states which are controlled uniquely by a single choice

(chosen on the basis of location) as a function of the rank of the choice, in order of frequency of being chosen. This number increases with rank ($r = 0.92^{**}$), at least for the first 8 choices. We do not know at what rank this tendency disappears. (The labor involved in producing the subjective totals prohibits the continuance of the graph.) However, sooner or later, an informant's choices are used only once or twice, at which time one can hardly assume that they "account for a state." We suspect that the degree of specialization remains high for at least 30 choices, for reasons given shortly.

Although specialization towards handling a single location increases with rank of choice, there is no accompanying increase in specialization of occupation, the level of which remains roughly constant over the top eight choices. Given that the number of triggers in a given occupation is of necessity smaller than that in a given state, it may be that specialized occupation choices occur much further down the list of choices.

As a rough guide, there is little sharing of function between the eight top choices of any informant. The list of occupations of triggers, selected on the basis of occupation, shows no overlap among the top eight choices of any informant. When overlap of location, selected on the basis of location, occurs, it is mainly of the most minor kind. An example of this is informant 42. His top choice controlled the Far West, Rocky Mountains, South West and Plains. His later choices also controlled large areas. Missouri occurred in each of the top 4 choices' areas of control. However, choice 1 had one occurrence of Missouri; choice 2, 3 occurrences; choice 3, 17; and choice 4, 2. Thus, despite the overlap, choice 3 is clearly handling Missouri. Apart from isolated cases of equal sharing of a state between two choices, the above example represents an extreme case of overlap. For the most part, overlap did

not occur.

The geographic areas controlled by an informant's high choices are of great interest. Figures 7 and 8 show the mean fractional amount of each state controlled by the top 8 choices: Fig. 7 when only choices made on the basis of location are considered, Fig. 8 for all reasons of choice. The total population of a state is defined as the number of triggers in that state. The figures possess two common features: the consistent large fraction of the Far Western states accounted for, and the almost nonexistent occurrence of local states (West Virginia, Pennsylvania, Ohio), the Mid-Eastern states, and Florida.

The latter tendency persists even when the top 35 choices (accounting for over half the U.S.) are used, although it is most marked for the local states (38% cover for W.Va., 46% for Penn., compared with 60% for the others, reason for choice being irrelevant). This, we believe, can be interpreted very simply: since an informant's knowledge of his local area is presumably more detailed, it follows that the choices for local triggers will tend to be more specialized -- a dentist in this county, a farmer in that county, and so on. The result is that each local choice will be used infrequently, as the number of triggers that a local choice can handle is by definition small.

This suggests, then, that informants do have at least two distinct networks: one of choices handling the U.S., and another, disjoint, set handling local triggers. (There are indications that a third set, handling the rest of the world also exists, but there are insufficient data on this point. See Fig. 2.) Jumping ahead somewhat in logical sequence, this statement is confirmed both in Section 10 and by the following. Let us compute the fraction (number of different choices on the basis of location • number of triggers)

in each state. Then the mean fraction for local states is 0.27 (s.d. 0.13) and for nonlocal states, 0.14 (s.d. 0.05). The difference is highly significant** and demonstrates that proportionally many more choices are used for location in local states.

7. The Top Choice

The choice used most frequently by informants has well-defined characteristics. Most (33) are male; less (23) are female. There was no tendency for males to choose a male top choice in preference to a female top choice; the same was true for female informants. Friends and acquaintances account for 37 of the top choices; 21 were family. This proportion is skewed towards use of family significantly** more than the overall probability in Table 2.

The top choice was used on average for location 58 times (s.d. 83); for occupation 26 times (s.d. 40); for race 3 times (s.d. 14, but a large maximum of 96); and for other reasons 38 times (s.d. 115, maximum 771). With the exception of the "other" category, skewed untypically high by one informant, this pattern of reasons again reflects those for the totality of choices.

Various characteristics of the informants accounted for some of the variation in the usage of their top choices. For example, female informants use their top choice significantly** more often than do males. More educated informants use their top choice significantly less* than less educated informants. Older people, and those from nonrural backgrounds tend* to choose females for their top choice. All informants from rural backgrounds (5) use family for their top choice**.

Male top choices are significantly* associated with informants who make

fewer choices in total (168 for male top choices, 268 for females). Male top choices are also used significantly* more for location (82) than are female top choices (30).

Finally, the most dramatic result, top choices among family are used much more frequently** on the basis of occupation (53) than are friends and relations so used (10).

8. Informants and Reasons for Their Choices

In this section we examine the number of times informants made choices based on the four reasons offered to them in the questionnaire, namely location, occupation, race/ethnicity, and "other."

On average, over all the informants, location was used as a reason 716 times (57% of the time) s.d. 318; occupation was used 476 times (38%) s.d. 265; race/ethnicity was used 21 times (2%) s.d. 34; and "other" was used 106 times (8%) s.d. 171. (The percentages sum to greater than unity as more than one reason could be used per trigger.) Thus, overriding reasons for choice are, first, location, and second, occupation.

Very little of the characteristics of the informants accounts for variation in the number of times they made choices based on any of the four reasons. An informant's education is positively correlated with the number of times he made choices based on location (0.22*), and the number of choices made on the basis of race/ethnicity (0.25*). It is also negatively correlated with the number of choices made on occupation (-0.25*). The only other characteristic to account for any variation was, curiously, that part-time housewives used race/ethnicity as a reason significantly* more than full-time and nonhousewives

(32 times per informant as against 15 on average).

In other words, the characteristics of informants do not affect the reason they make their choices very strongly. The following two sections examine what *does* affect the choice-making procedure.

9. Triggers and Choices

If characteristics of the informants fail to account for much variation in reason for choice, then perhaps characteristics of the triggers do. The reader will appreciate that there are a great many intercomparisons possible: there are five characteristics of each trigger (sex, size of location, occupation, race/ethnicity, and location itself); effectively two for each choice (sex and family or friend/acquaintance); and four reasons for choice. The complexity of a full 3-way comparison is prohibitive. Instead, we did pairwise comparisons: those involving the reason for choice (dealt with in the next section); and those that didn't, which are dealt with here.

First, size of the trigger's town or city had no effect on the type of choice. However, all other details of the trigger had a significant effect on the type of choice selected by informants. Table 5 shows these results.

Male choices tend to be made no matter what the sex of the trigger, although the probability of making a female choice goes up for female triggers. The probability of making a female choice for a female trigger is 46%, compared with 29% for a male trigger (irrespective of the sex of the informant). This is in contrast to the suggestion by Travers and Milgram (1969) that choices tend to be of the same sex as the target person; in fact, the results suggest the sex of the target persons to be almost irrelevant in selecting the sex of

for	Number of times male choice used	Number of times female choice used	χ^2(1 d.f.)	Number of times friend/acqu. used	Number of times family used	χ^2(1 d.f.)
male trigger	25432(+ ,71)	10485(- ,29)	7300**	28954(+ ,79)	7764(- ,21)	
female trigger	17574(- ,54)	14900(+ ,46)		25708(- ,78)	7369(+ ,22)	13**
low occupation trigger	24100(- ,60)	16179(+ ,40)	425**	31569(- ,77)	9443(+ ,23)	110**
high occupation trigger	21037(+ ,67)	10206(- ,33)		25823(+ ,80)	6378(- ,20)	
American trigger	30741(- ,62)	18454(+ ,38)	26**	39181(- ,78)	10956(+ ,22)	5.5*
ethnic trigger	14396(+ ,64)	7931(- ,36)		18211(+ ,79)	4865(- ,21)	
near trigger	3445(- ,61)	2219(+ ,39)	14**	Not significant		
far trigger	41692(+ ,63)	24166(- ,37)		Not significant		

TABLE 5. Contingency tables between types of trigger and types of choice. Entries are the number of times (over all informants) that a particular combination occurred. A (+) or (-) after an entry indicates the direction from the expected frequency of that cell. The number following this sign is the chance of making a choice in their category. Low and high occupations are defined as ≤ 50, >50 on the Duncan-Reiss scale respectively; "ethnic" triggers are the five categories defined in Section 2; near triggers are those in West Virginia, Ohio, and Pennsylvania; far triggers otherwise.

choice. The sex of the choice varies with the occupation of the trigger (given that male choices are likely to be made anyway). Male choices are more likely than expected to be made for high occupation triggers (67%) and less likely than expected for low occupation triggers (60%). Males choices are also more likely than expected to be used for ethnic triggers (64%), as against non-ethnic triggers (62%). Male choices are less likely than expected to be used for "near" triggers (those residing in West Virginia, Pennsylvania and Ohio) (61%) than for "far" triggers (those residing elsewhere) (63%).

Friends or acquaintances (hereafter, friends) tend to be used more often than expected for male triggers (79%) than for female triggers (78%). This apparently small difference is, nevertheless, highly significant, because of the size of the entries in the tables. Friends are used more often than expected for high occupation triggers (80%) than for low occupation triggers (77%). Finally, friends are used more often than expected for ethnic triggers (79%) than for nonethnic triggers (78%). Whether a trigger was "near" or "far" had no effect on whether a family member or a friend was chosen.

So far, we have been discussing the number of choices made, or the probability a particular type of choice would be made. It is also of interest to consider how many different choices an informant makes for triggers in a given category, which we shall call "the variety of choices." Consider Table 6.

The most obvious fact about Table 6 is its remarkable similarity to the left side of Table 5. There is, after all, no reason why this should be the case, since the two tables present two entirely different sets of measurements. Still, the fact remains that male triggers generate a much wider variety of male choices than female choices, whereas female triggers generate a more even variety of male and female choices. Similarly, the percentage variety of

for	Number of different male choices used	Number of different female choices used	Number of different friends/acqu. used	Number of different family used
male triggers	101.1 (69%)	44.4 (31%)	133.5 (88%)	18.2 (12%)
female triggers	77.0 (58%)	56.4 (42%)	120.1 (87%)	18.0 (13%)
occupations ≤ 20	55.2 (59%)	38.9 (41%)	81.7 (84%)	15.5 (16%)
21 ≤ occupations ≤ 40	43.5 (66%)	22.9 (34%)	57.9 (84%)	10.7 (16%)
41 ≤ occupations ≤ 60	92.1 (63%)	55.2 (37%)	78.7 (86%)	13.3 (14%)
61 ≤ occupations ≤ 80	59.9 (64%)	33.6 (36%)	83.5 (86%)	13.6 (14%)
81 ≤ occupations	42.0 (71%)	16.9 (29%)	54.1 (87%)	8.1 (13%)
American triggers	106.7 (63%)	62.1 (37%)	154.5 (88%)	20.9 (12%)
Black triggers	44.7 (65%)	24.0 (35%)	59.7 (84%)	11.1 (16%)
Spanish triggers	28.9 (64%)	16.4 (36%)	38.5 (83%)	8.0 (17%)
Italian triggers	21.0 (59%)	14.3 (41%)	29.8 (83%)	6.2 (17%)
Slavic triggers	17.8 (64%)	10.2 (36%)	23.3 (81%)	5.3 (19%)
Oriental triggers	14.3 (65%)	7.8 (35%)	19.2 (82%)	4.1 (18%)

TABLE 6. Table showing the mean number of <u>different</u> choices used in each category, i.e., an indication of the variety of choices used. The percentage figure represents 'the fractional amount of variety' and is directly comparable with those in Table 5.

male choices tends to increase with the occupation level of the trigger.

Some differences between Tables 5 and 6 do occur. Specifically, the fractional variety of friends (Table 6) is distinctly larger than the fractional number of friends used (Table 5), although the pattern of increase and decrease with category of trigger persists. This suggests family choices are selected for rather more specialized triggers than are friends.

There is one major difference between Tables 5 and 6. In Table 5 the probability of choosing a friend for an "ethnic" trigger was higher than expected. In Table 6 the percentage variety of friends used for such triggers is lower than expected. This again suggests that specialized friends are used for specialized triggers.

Are these changes in variety a function of other features of the triggers? In general, the answer is no. Suppose the triggers are split by two categories, occupation level and sex, for example. Define the fractional variety in each cell of such a table as the mean number of different choices used for triggers in this pair of categories divided by the number of triggers in this pair of categories. Thus, a fractional variety of 1 would indicate that a different choice was used for each of the triggers in the pair of categories. Only two results of merit turned up. When the triggers are split by occupation and sex, there is a weak tendency (except for housewife triggers) for a greater fractional variety for female triggers than for male triggers. This indicates that, although housewives account for 25% of the female triggers, the variety associated with them is very low (0.33) and this tends to lower the total variety for female triggers.

A much stronger association appears when triggers are split by occupation

level (in ten percentiles, to avoid too low numbers in each cell) and race/ethnicity. The fractional variety for nonethnic triggers over all occupation levels is low (about 0.5), whereas for all ethnic triggers the fractional variety averages about 0.9. Note, however, that these high values disappear if the triggers are split by any category other than occupation. In other words, so long as the trigger is "ethnic" informants tended to use a different choice for each such trigger.

10. Triggers and Reasons

Hitherto little has appeared to account for how and why particular types of choices are made by informants. If the reasons which informants give for their choices are examined in conjunction with characteristics of the triggers, then a pattern begins to emerge.

Table 7 shows contingency tables between triggers and the various reasons for choices, split into various categories of trigger. Nearly all splits give significant results. Female triggers are significantly** more likely than expected to be chosen on the basis of location (and more likely than male triggers, incidentally), and correspondingly, much less likely** to be chosen on the basis of occupation. Location is also used as a reason more frequently** for triggers in large towns or cities than for small towns, and the reverse is true for occupation as a reason**.

Splitting triggers into their ethnic groups yields some interesting results. Location is uniformly less likely** to be used as a reason for triggers who are in any way ethnic; however, when used as a reason, occupation shows no such tendency. Not surprisingly, race/ethnicity is used more frequently** as

	Number of times reason							
for	Location not used	Location used	Occupation not used	Occupation used	Race/ ethnicity not used	Race/ ethnicity used	Other not used	Other used
male triggers	15939(+,44)	20503(-,56)	22184(-,61)	14258(+,39)	not		33379(,.92)	3063(,.0
female triggers	13348(-,41)	19474(+,59)	20757(+,63)	12065(-,37)	significant		30257(,.92)	2565(,.0
	$\chi^2 = 66^{**}$ 1 d.f.		$\chi^2 = 41^{**}$ 1 d.f.				$\chi^2 = 8.1^{**}$ 1 d.f.	
small town	16306(+,45)	20106(-,55)	22110(-,61)	14302(+,39)	35815(-,.98)	597(+,.02)	33264(-,.91)	3148(+,.
large town	11939(-,39)	18875(+,61)	19561(+,63)	11253(-,37)	30371(+,.99)	443(-,.01)	28721(+,.93)	2093(-,.
	$\chi^2 = 249^{**}$ 1 d.f.		$\chi^2 = 54^{**}$ 1 d.f.		$\chi^2 = 4.5^{*}$ 1 d.f.		$\chi^2 = 80^{**}$ 1 d.f.	
American	20362(-,41)	29395(+,59)			49583(+,1)	174(,0)	45560(,.92)	4197(,.0
Black	4191(+,47)	4693(-,53)			8279(-,.93)	605(+,.07)	8172(,.92)	712(,.0
Spanish	2020(+,46)	2387(-,54)	not		4274(-,.97)	133(+,.03)	4073(,.92)	334(,.0
Italian	1296(+,43)	1749(-,57)			2993(-,.98)	52(+,.02)	2854(+,.94)	191(-,.
Slavic	1041(+,43)	1372(-,57)	significant		2377(,.99)	36(,.01)	2248(+,.93)	165(-,.
Oriental	942(+,47)	1042(-,53)			1846(-,.93)	138(+,.07)	1812(-,.91)	172(+,.
	$\chi^2 = 170^{**}$ 5 d.f.				$\chi^2 = 356^{**}$ 5 d.f.		$\chi^2 = 28^{**}$ 5 d.f.	

TABLE 7. Contingency tables for reasons and trigger characteristics. Entries are total number of times a given combination occurred, followed by a sign and percentage as in Table 5. No sign indicates no deviation from norm to 2 digits.

Mean occupation level when location used as reason	43
Mean occupation level when location not used as reason	51
$t = 42^{**}$, 72586 d.f.	
Mean occupation level when occupation used as reason	53
Mean occupation level when occupation not used as reason	42
$t = 58^{**}$, 72586 d.f.	
Mean occupation level when race/ethnicity used as reason	46
Mean occupation level when race/ethnicity not used as reason	46
$t = 0.25$, 72586 d.f., not significant	
Mean occupation level when "other" used as reason	42
Mean occupation level when "other" not used as reason	46
$t = 11.7^{**}$, 72586 d.f.	

TABLE 8. T-tests between the occupational level of triggers when divided into subsets by the reason for choice.

a reason for choice on ethnic triggers. However, this tendency is not uniformly distributed between the various ethnic groups. If χ^2 is recomputed among the five ethnic groups, one finds that blacks and Oriental triggers are significantly ($\chi^2 = 246^{**}$, 4 d.f.) more likely to produce a choice on the basis of race/ethnicity than are Spanish, Italian, or Slavic triggers.

We believe the reason to be as follows: Informants can easily distinguish black triggers (coded by a 'B' in the questionnaire), and also Orientals, by their names. However, Spanish, Italian or Slavic triggers are not obviously ethnic -- they may have acquired the name by marriage, may be third generation, etc. Thus race/ethnicity is a "safe" reason only when the trigger is obviously ethnic; hence the high probabilities for black and Oriental triggers.

Table 8 shows the equivalent split of occupation level by the four reasons for choice. The mean occupational level of triggers was significantly** higher when occupation was used as a reason than when not; significantly lower** when location was used as a reason than when not; significantly lower when "other" was used as a reason than when not; and, finally, race/ethnicity used as a reason produced no significant variation in occupational level. The indication, then is clear. The higher the occupational level of the trigger, the less likely is location to be used as a trigger, and the more likely is occupation to be used as a trigger. We defer discussion of the "other" reason until later.

This finding brings to mind the well known phrase "my man in...."[3] where the blank is filled in with some location. If high occupation triggers are more likely to generate choices on the basis of occupation, this suggests that informants may have "their ____" where the blank is filled by a high status occupation (dentist, architect, etc.). Whether this is true, and whether informants have "their man in...." can be tested, although precise definitions

of these terms are problematic. Several versions will be given as we proceed.

To begin with, we saw in Section 6 that on virtually no occasion was an entire state covered by only one choice. One could, therefore, ask two distinct questions: first, how many different choices are used for triggers in a given state when location is given as the basis for choice, i.e., what is the variety of choices? Expressed as a fraction of the number of times location is used as a reason for that location, the variety has an average of about 0.35 and declines weakly with distance from Morgantown, West Virginia. (Distance is defined to the center of the state in question, and measured, lacking any more accurate measures, in cm. on a map. Alaska and Hawaii are purposely omitted in what follows.) This suggests that one tends to have a "man in" states further from home.

Perhaps a more relevant question is: how much of a given state is accounted for by the best single choice, when it is chosen on the basis of location? It turns out that, on average, an informant's "best person" in a state handles 69% (s.d. 9%) of the occurrences in the questionnaire, when choices are made for these occurrences on the basis of location. This figure is both surprisingly high and surprisingly uniform, and suggests strongly that the concept of "a man in...." is indeed a valid one for the small-world problem. But how many states does an informant have "a man in," and where are they?

If one defines a cutoff for "having one's man in" of 67% of all occurrences on the basis of location, then the mean number of states an informant has "his man in" is 24.9 (s.d. 8.8). Put another way, 49% (s.d. 17%) of all states in the U.S. are controlled by a single person on the basis of location.

The distribution of the control is not uniform over the U.S. Figure 9 shows the amount of the occurrences of a single state controlled by one person

on the basis of location, as a function of the distance of that state from Morgantown. Although the correlation (0.26) is not significant, the impression one gets is of low amounts by near states (West Virginia, Pennsylvania, Ohio) and almost uniform amounts for other states, independent of distance, as indicated by the superimposed lines. California is a notable exception.

But is the probability of choosing on the basis of location constant between states? The answer is, of course, no. There is a weak tendency ($r = -0.34^{*}$) to choose on the basis of location less, the further the state is from Morgantown. In fairness, most of this correlation is produced by the very high probabilities of using location for nearby triggers (76% for triggers in West Virginia, for example). These high probabilities act as further confirmation of the hypothesis in Section 6, namely a great many specialized choices form part of an informant's local network. For nearby triggers, then, one would expect location to be the overriding reason for choice.

Having seen that the concept of "one's man in" is valid, we return to the parallel concept of "one's___" (high status occupation). Again, one can examine this first by the variety of choices for a given occupation. The fractional variety of choices per occupation level (defined in analogous fashion to that for location) declines strongly ($r = -0.41^{**}$) as occupation level increases. In other words, one uses fewer choices for triggers of a given occupation as the occupation level goes up.

One can also examine the probability of using occupation as a reason, as a function of the occupational level of the trigger. (One might suspect that this depends on some of the characteristics of the informant. However, this was not the case for any of the informant variables we collected. Unfortunately,

we did not have directly comparable information on informants' and triggers' occupations.) As Figure 10 shows, the probability of using occupation as a reason increases strongly ($r = 0.68^{**}$) with the occupation level of the trigger. (The peak at level 42 is 'photographic processor.' Apparently most informants have "their photographic processor.")

So striking was this finding that attempts were made to explain the residual variance by location of the trigger, along the lines suggested above. The triggers were subdivided into "near" and "far," and the equivalent curve to Fig. 10 was produced for each subgroup. However, there is no significant difference (less than 2%, in fact) in the resulting correlations, suggesting that location does not bear directly on the decision to choose on the basis of location.

Further confirmation for Fig. 10 comes from Travers and Milgram (1969). The target person in their small-world experiment was a stockbroker (85 on the Duncan-Reiss scale). The best-fit line, and, coincidentally, the data, give a probability of 0.5 of making a choice on the basis of occupation when the trigger is at occupation level 85. Now Travers and Milgram report that 60.7% of the links generated from the "stockholder group" of starters, and 31.8% of those generated from the "random group" of starters reported employment connected with financial matters. Then each person is a choice of someone further back the chain, with a stockbroker as target. Hence we would predict that 50% of these choices should be made with occupation as the reason, and therefore have occupations connected with finance. Using the numbers in the two groups given by Travers and Milgram, a value of 48% or 46% is produced depending on whether complete or incomplete chains are assumed (their text is not specific).

These empirical values given by Travers and Milgram are in excellent agreement with our predictions.

Similar curves to Fig. 10 were drawn for the probability of using location, race, or "other" as a reason for choosing a trigger of a given occupation level. The resulting correlations between the reason and occupation level were: with location, -0.59^{**}; with race, 0.02; and with "other," -0.53^{**}. Hence the probability of using location as a reason is very high for low status occupations; the probability of using "other" as a reason is also at its highest for low status occupations, but its numerical value is still very low; and, as found previously, race/ethnicity appears to be independent of occupation level.

The conclusion to be drawn from this is that, overwhelmingly, the predominant feature of the trigger which controls which type of choice is made is the status of the trigger's occupation: if high, the choice will probably be made on the basis of occupation; if low, on location. Put another way, one does have "one's dentist" but not "one's ditchdigger."

A note is in order about choices where "other" is given as the reason for making the choice. As far as predicting the reason for a choice is concerned, "other" appears, from our data, like a weaker variety of location. The informants had "their man in" a variety of states, providing that the status of the trigger was high enough. In such cases, location was the reason for the choice. For lower status triggers, the same "man in" was used for the choice, but the reason given changed to "other." We believe that this effect occurs because informants do not, for some reason, wish to associate "their man in" some location with too low a status trigger, and hence invent -- in their own minds -- another reason for making such a choice. Lacking a name for this effect, we term it the G effect, for historial reasons.

11. Accuracy of Informants

As noted in Section 2, informants were asked, some months after they completed the questionnaire, how well they recalled their answers. Although posed as a test of recall, we prefer to think of this as a test of how well informants understand the network in which they are embedded. This is a valid comparison provided (a) that an informant's world network is changing at a sufficiently slow rate and (b) that the picture of the network emerging from the data is a meaningful one. With these assumptions, it follows that if an informant understood his network at the time he completed the questionnaire, he still understood it when asked about it later. Furthermore, what he understood was essentially the same as when he provided the data about it.

Therefore it seems relevant to enquire how well his understanding of the network agrees with what the network actually looks like. As a simple measure, informants estimated how many different choices they had used. The answers obtained (from 40 informants) are summarized in Figure 11. In only one case did an informant over-estimate, and then by only 1%; the informant concerned was associated with the researchers, a trait which clearly improved informant's powers of estimation.

In general, informants guessed that they had used an average of about 50 choices. This is about one-third the actual total for the group concerned. In no case was a guess larger than 200 made. There is a positive correlation of 0.33* between correct and estimated number of choices, which at least indicates that guesses of network size do tend

to increase with network size. However, the consistent and blatant underestimates suggest that informants have no real idea of the size of their networks.

There is a plausible reason for this. As was seen in Section 6, choices which handle local triggers are not frequently used in the questionnaire; yet in day-to-day existence these choices are used continually. We suspect that the top 35 choices, although they do handle half the U.S. for most informants, are so infrequently used in practice that informants simply forget about them when answering our question.

Indications of this also come from the responses to "who are the three choices you used the most?" (The similarity between this question and the traditional sociometric question is deliberate.) Surely, we felt, informants ought to be able to name, say, their most frequently used choice. After all, that choice accounted, on average, for 10% of the triggers. Unfortunately, this was not the case.

On average, informants only guessed the name of their most frequently used choice 43% (s.d. 50%) at the time. A similar pattern emerges if one examines how many of the top three choices were correctly identified, irrespective of order. (This is the $\hat{A}$ score of Killworth and Bernard, 1976). The mean score was 1.4, s.d. 1, so that only one-third of the most frequently use network is identified correctly; we suspect that this is for the same reason as given above.

It was possible that one or several of the informants' characteristics could account for the variation in accuracy, although this seemed unlikely from our previous results (Bernard and Killworth, 1977). However, it

turned out that one characteristic did account for some of the variation of accuracy: if the informant was in part-time employment, his accuracy was low; if in full-time employment or unemployed (for whatever reason), his accuracy was high. As confirmation, this also showed up in the "housewife" subcategory of "part-time housewife." We are at a loss to explain this, unless it reflects a genuine difficulty of the part-time employed to handle different networks-- work, home, friends-- in what must be rapid succession in day-to-day life.

By an unfortunate oversight, the length of time between an informant completing the questionnaire and his being asked to recall his top three choices was not recorded. Hence we can not compute whether there is a deterioration in recall with time. However, in many other experiments (Killworth and Bernard, 1976; Bernard and Killworth, 1977) we have found that informant accuracy in recall of their communication (in both open and closed networks) was unreliable even when tested just after the period of communication. Given also that we should expect an informant's global network to be fairly stable with time, there is no a priori reason to suppose that informant inaccuracy in this experiment is a function of the relatively short (1 - 4 months) time between questionnaire and recall.

12. Conclusions

This experiment examines informants' cognition, or their guesses about how they would behave under specified conditions. Furthermore, the specified behavior (the small-world problem) never occurs in the real world, except when social scientists create it. This creates obvious limitations, so

far as generalizing from these data is concerned, but these limitations are no more or less stringent than those which must be imposed on any other form of sociometric research. Ultimately, we want to understand social structure, here defined as the pattern of overlap both between informants' cognitive networks, and between their behavioral networks. The reverse small-world experiment provides a very rich source of data on people's cognitive networks.

The experimental design, with hindsight, possessed some drawbacks. Many more "local" triggers should have been included. Informants frequently told us that they "never got a chance to use" many of their acquaintances who lived nearby. In future experiments all important personal data about the choices (their occupation, precise location, age, etc.) should be collected. As it stands now, we can't tell whether someone's "man in Idaho" actually lives in Idaho. In future experiments, larger and more representative samples of informants should be used.

Perhaps the largest drawback is caused by the size of the data. Ideally, one would like to create a mean "cognitive map" of all the residents of the U.S. for which the first 1167 triggers serve as proxy. What in general makes some triggers seem like others to informants? There is an obvious tool to answer this question: multidimensional scaling (MDS). It is quite straightforward to generate a matrix of similarity coefficients between triggers (a count of the number of times the same choice is used for each pair of triggers will do). Then an MDS would yield subgroups of the triggers which are perceived as similar, and conclusions could be drawn from these groupings. Unfortunately, the size of the matrix concerned (over 10^6 entries) prohibits this analysis on present computers.

Were it not for Milgram's original experiments, this research would have to apply only to cognitive processes. However, the strong similarity between predictions from some of our results and the results of Travers and Milgram (1969) does suggest that the reverse small-world experiment is an adequate proxy for behavior. Two points bear directly on the comparison between the experiments. First, in ours informants were not required to mail anything to their choices. In many cases informants reported that they were unsure of the current address of some of their choices. The instructions for the experiment only specified that informants a) know their choices well enough to ask them to participate in a small-world chain; and b) that they be able to contact their choices if necessary. We suspect that there would be a great tendency to restrict the number of choices used in an experiment where actual contact with choices was involved. This restriction would favor the use of choices people know best. Furthermore, in our experiment, some informants tried to "use as many different names as possible" in spite of repeated instructions that this was not necessary.

Another feature of the reverse small-world experiment is that the most frequent choices did not necessarily appear early in the responses to the questionnaire. As an indication, on average 12.3 (s.d. 3) of the 26 most frequently used choices were not among the first 20 choices triggered. There are two possible reasons for this. First, the order of the triggers may not have been sufficiently random. However, recall that reversing the order of the triggers produces no significant difference. This, therefore, suggests that the second explanation is more likely to be correct. This is that people get accustomed to the experiment as they get deeper into it. In fact, several informants said that the first 10 pages or so of the questionnaire helped to

get them "warmed up" to the task of selecting appropriate choices. Many persons said that the experimental task helped them to "dredge up" people they hand't thought about in years. Of course, this was precisely the point of the technique.

The result of this is that "more appropriate" choices may tend only to be triggered long after the beginning of the questionnaire. If this is true, it suggests that single realizations of small-world experiments may be generating a behaviorally biassed picture of people's cognitive network. This, of course, is the direct opposite of the possibility, noted above, that the reverse small-world experiment may be generating a cognitively biassed picture of the behavioral world. As we have shown elsewhere (Killworth and Bernard, 1976), there are probably two quite distinct "social structures" based on cognition about one's place in a network and one's actions within that network. Exactly how the discrepancies between these two networks are handled by individuals needs urgent research.

To sum up, our main findings are as follows:

1. A mean of 210 choices control the world, but this number is an underestimate. Only 35 choices control half the world, however.

2. There appear to be three distinct, and mainly nonoverlapping, networks controlling U.S., local, and foreign areas.

3. Choices are mainly friends and acquaintances, with strong cleavages by sex.

4. Location is the usual reason for choices, with occupation second most popular.

5. The decision as to which choice is made depends primarily on the occupation of the trigger, and secondly on the distance (near/far) from Morgantown.

6. Half the states in the U.S. are controlled by a single person when choices are made on the basis of location.

7. As found in previous experiments, the accuracy of informants in estimating their networks is low.

FOOTNOTES

[1] It is painfully obvious that the experiment required to produce the exhaustive network would require about 20 hours of work by each informant.

[2] Single asterisks denote significance at the 5% level; double at the 1% level or better.

[3] We were concerned about the connotation of the phrase "your man in Idaho." Aside from the fact that this rather sexist phrase is a common expression in English, it also turns out that the people who handle a particular state are males much more than they are females (62% vs. 38%). This appears to be an extension of the finding that more male choices are made anyway (64% vs. 36%).

REFERENCES CITED

Bernard, H. Russell and Peter D. Killworth
1977 Informant Accuracy in Social Network Data II. Human Communication Research, in press.

Bochner, S., et al
1976 Communication Patterns in an International Student Dormitory: Modification of the Small World Method. Journal of Applied Social Psychology 6(3):275-290.

Duncan, O.D.
1961 Socioeconomic Index Scores. In Albert J. Reiss, Jr., Occupations and Social Status. New York: Free Press.

Gurevitch, M.
1961 The Social Structure of Acquaintanceship Networks. Ph.D. dissertation. MIT, Cambridge.

Hunter, John E., and R. Lance Shotland
1974 Treating Data Collected by the "Small World" Method as a Markov Process. Social Forces 52(3):321-332.

Killworth, Peter D., and H. Russell Bernard
1976a Informant Accuracy in Social Network Data. Human Organization 35(3):269-286.

1976b A Comparison of Structure in Behavioral and Cognitive Social Network Data. Technical Report #KB-113-76. Office of Naval Research, Code 452, Arlington, Virginia.

Korte C., and S. Milgram
1970 Acquaintanceship Networks Between Racial Groups: Application of the Small World Method. Journal of Personality and Social Psychology 15(7):101-118.

Lundberg, Craig C.
1975 Patterns of Acquaintanceship in Society and Complex Organization: A Comparative Study of the Small World Problem. Pacific Sociological Review 18(2):206-222.

Milgram, Stanley
1967 The Small World Problem. Psychology Today 1(1):61-67.

Travers, J., and S. Milgram
1969 An Experimental Study of the Small World Problem. Sociometry 32(2): 425-443.

	NAME	LOCATION	OCCUPATION	RACE OR ETHNICITY	NAME	RELATIONSHIP	OTHER
542 (B)	Gomes, Wallace	Corbin, KY	Delivery man				
543	Kempner, Jill	Pittsburgh, PA	Shipping clerk				
544	Diaz, Ban	Los Angeles, CA	Mover				
545	Hopkins, Barbie	Pontiac, IL	Lawyer				
546 (B)	Solomen, Mimi	Aquila, AZ	Chemical Engineer				
547	Blukin, Marvin	Winchester, TN	Fountain manager				
548	Kessler, Matthew	Harrisburg, PA	Office worker				
549 (B)	Norel, Patty	Ainsworth, NE	Housewife				
550	Goltz, Sara J.	Huntington, WV	Claims examiner				
551	Horne, Joseph	Kawanee, IL	Gluer				
552	Kent, Frederic	Gordyce, AR	Symphony conductor				
553	Dickens, Sue A.	Las Vegas, NV	Housewife				
554	Sontz, Elvis	Denver, CO	Maintenance man				
555 (B)	Numberg, Newson	San Jose, CA	Buyer				
556	Pech, Melvin	Altoona, PA	Dentist				
557	Birbaum, June	Albany, NY	Professor				
558	McGuire, Tom	Bassett, NB	Office manager				
559	Digregorio, Mary	Rolling Fork, MS	Practical nurse				
560	Evanoff, Clark	El Centro, CA	Farm service worker				
561	Fuentes, Rosa	Portland, OR	Ballerina				
562	Paz, Ana	Dallas, TX	Midwife				

FIGURE 1. A specimen page of the questionnaire.

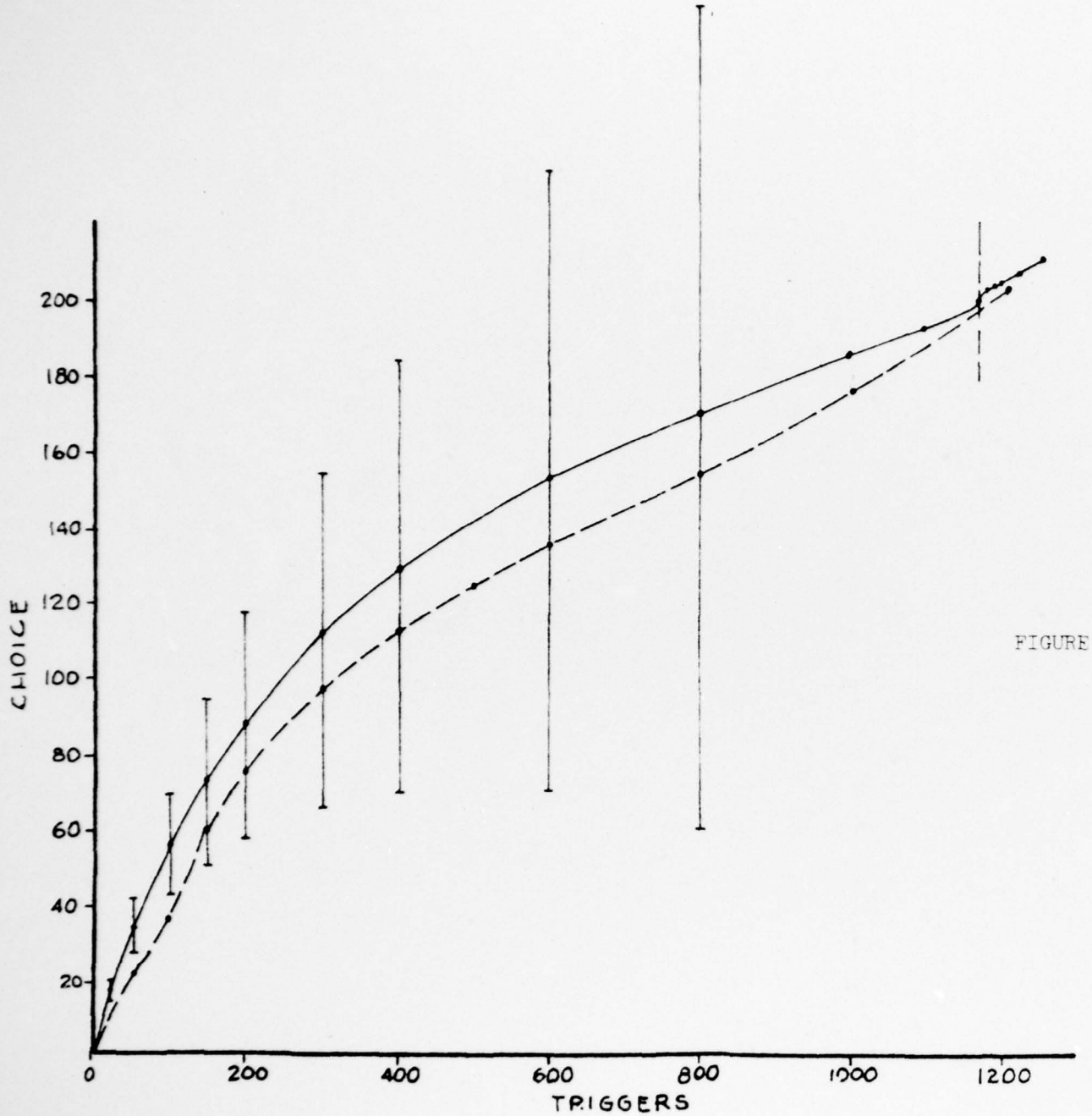


FIGURE 2. Cumulative histogram of the mean number of different triggers accounting for a given number of triggers. Error bars show one standard deviation. The change in gradient at trigger 1168 represents the beginning of the foreign triggers. The dashed line shows the same histogram with the triggers in reverse order, as a test of randomness.

FIGURE 3. Mean percentage of the world handled by the 20 most used choices.

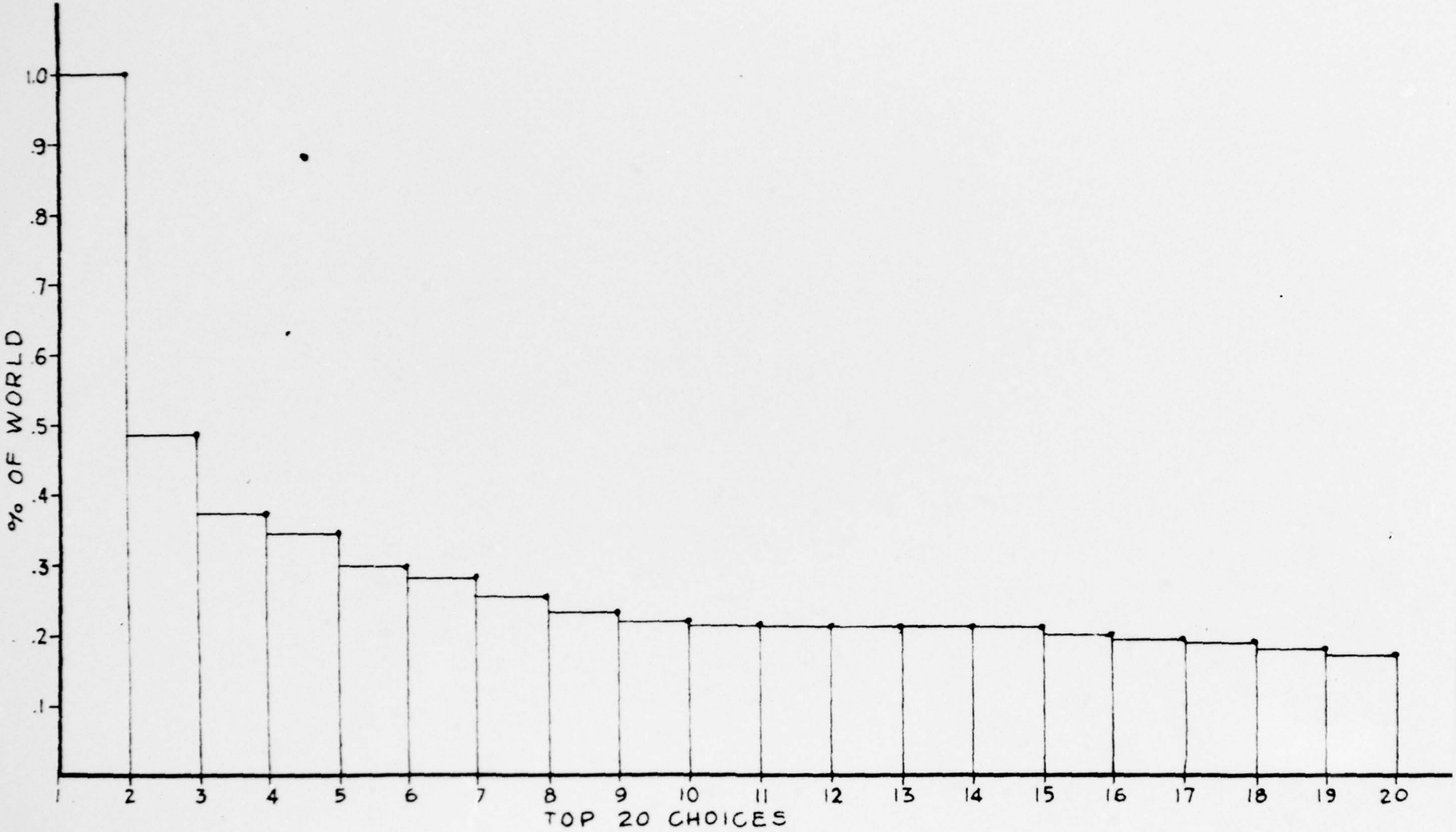


FIGURE 4. Number of choices made by informants.

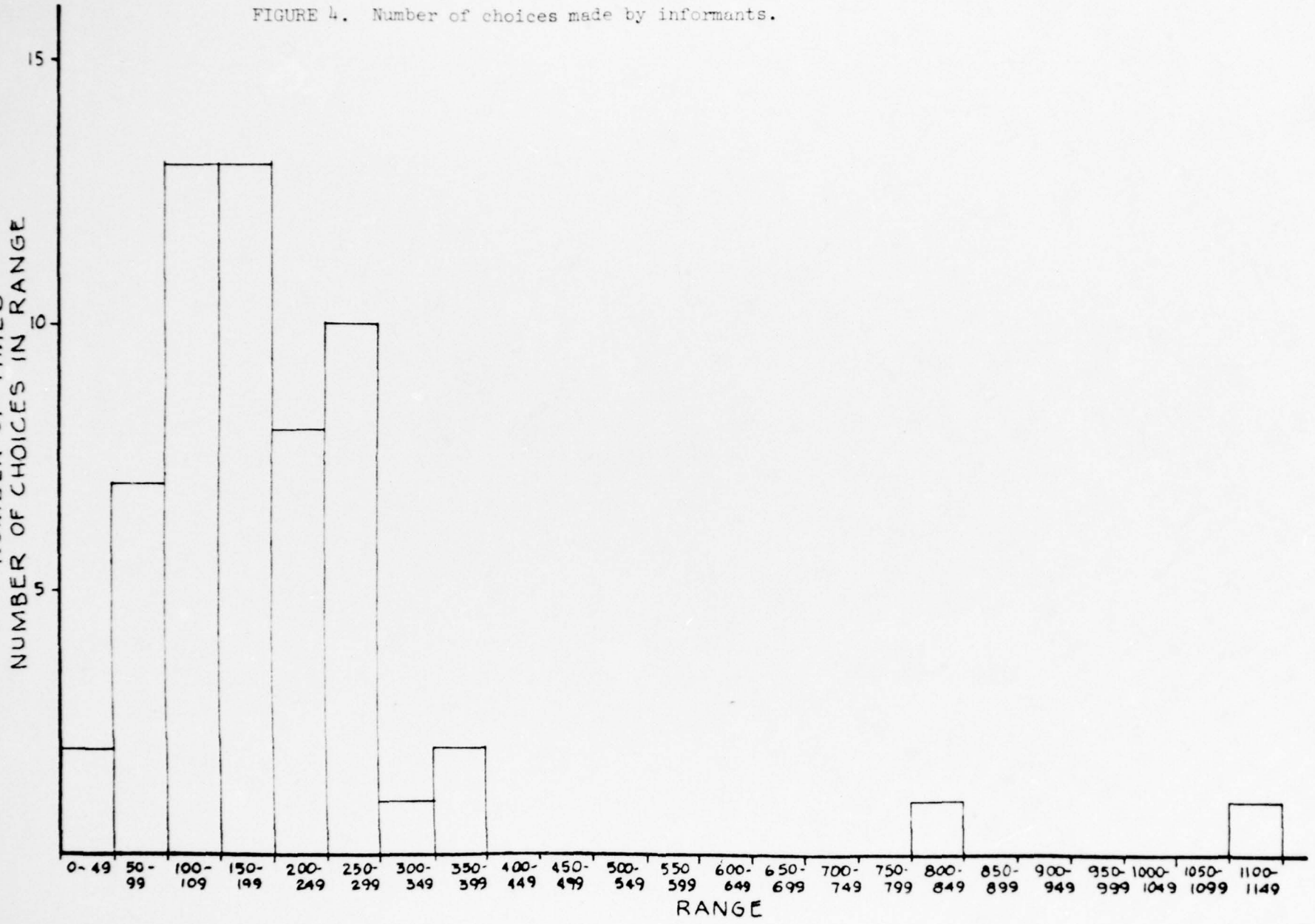


FIGURE 5. The U.S. divided into blocs as defined by the U.S. Office of Education.

FIGURE 6. Number of states which, when chosen on location, are only single-state for that choice. X-axis is the choice number, from 1 (most used choice) to 8th most used choice.

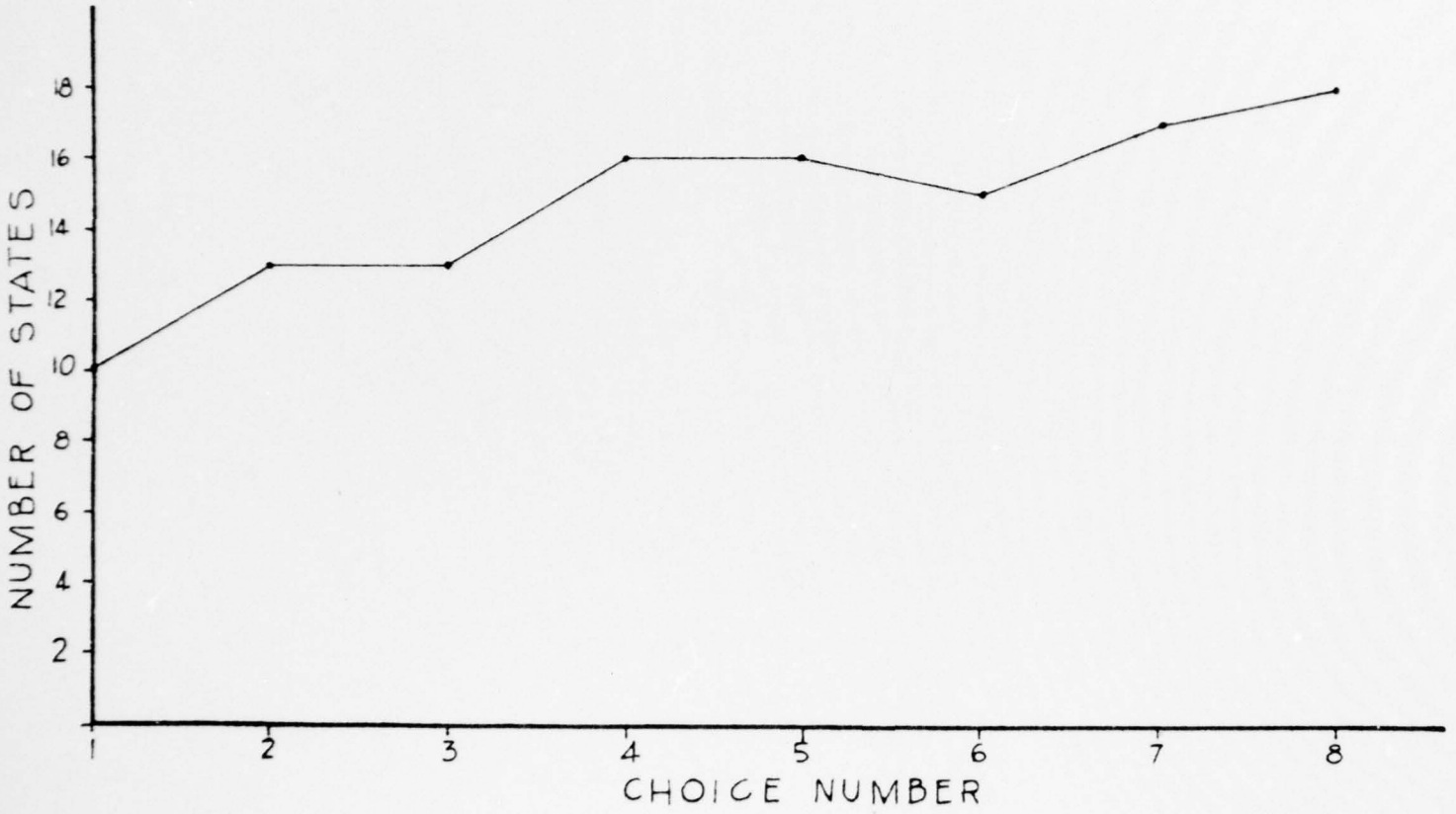


FIGURE 7. Percentage of each state handled (with location as reason) by top 8 choices.

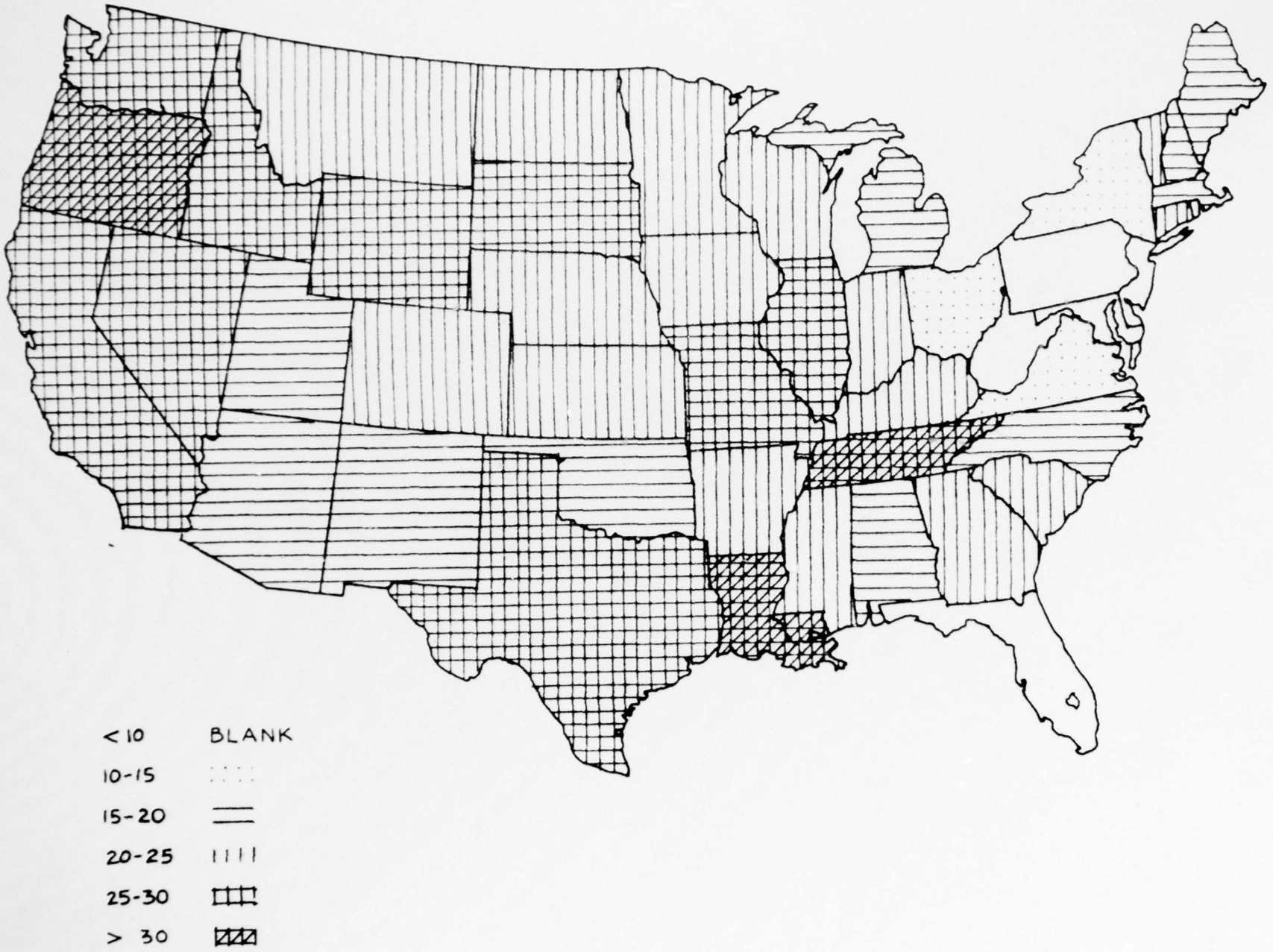
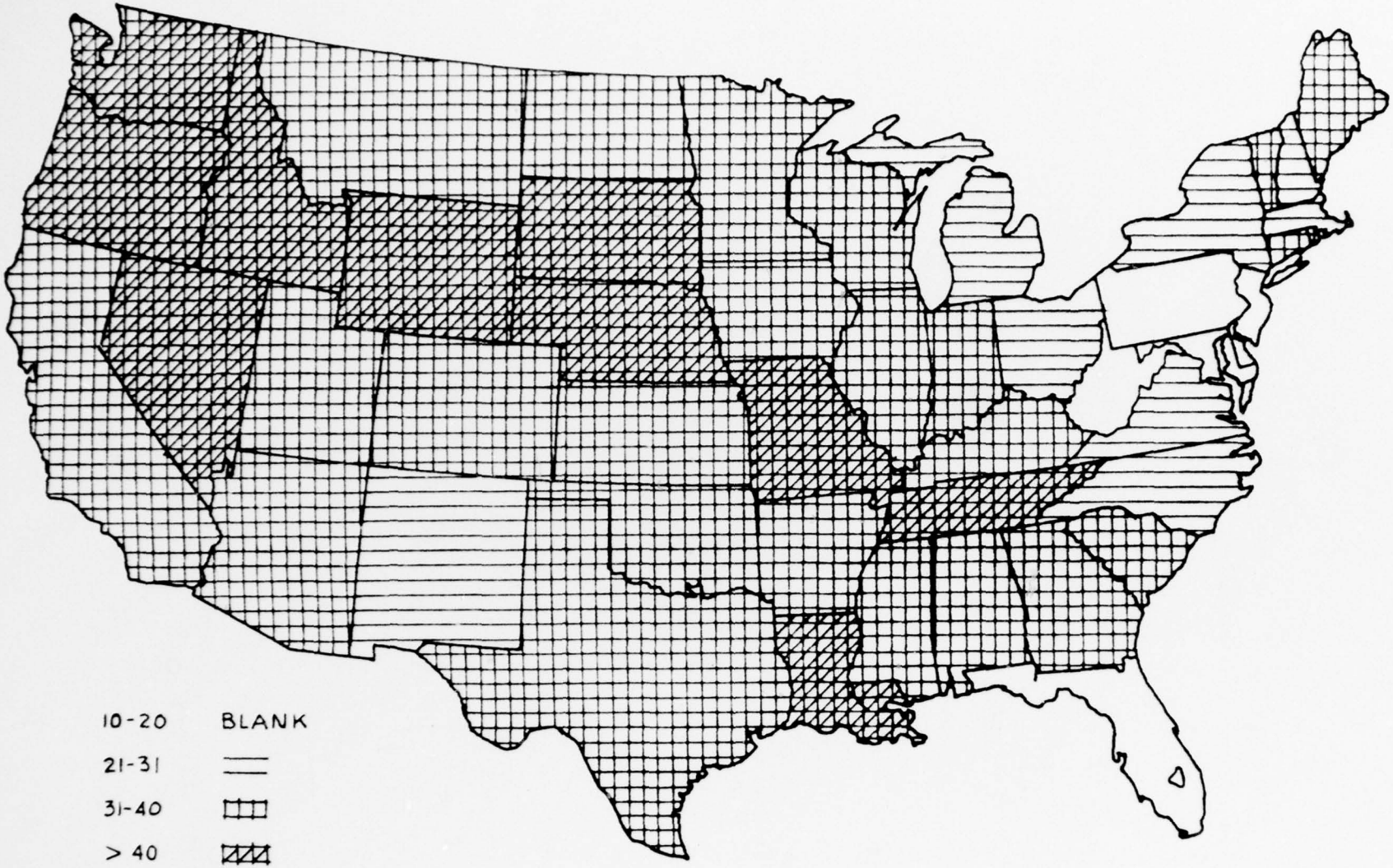


FIGURE 8. Percentage of each state handled, irrespective of reason, by top 8 choices.



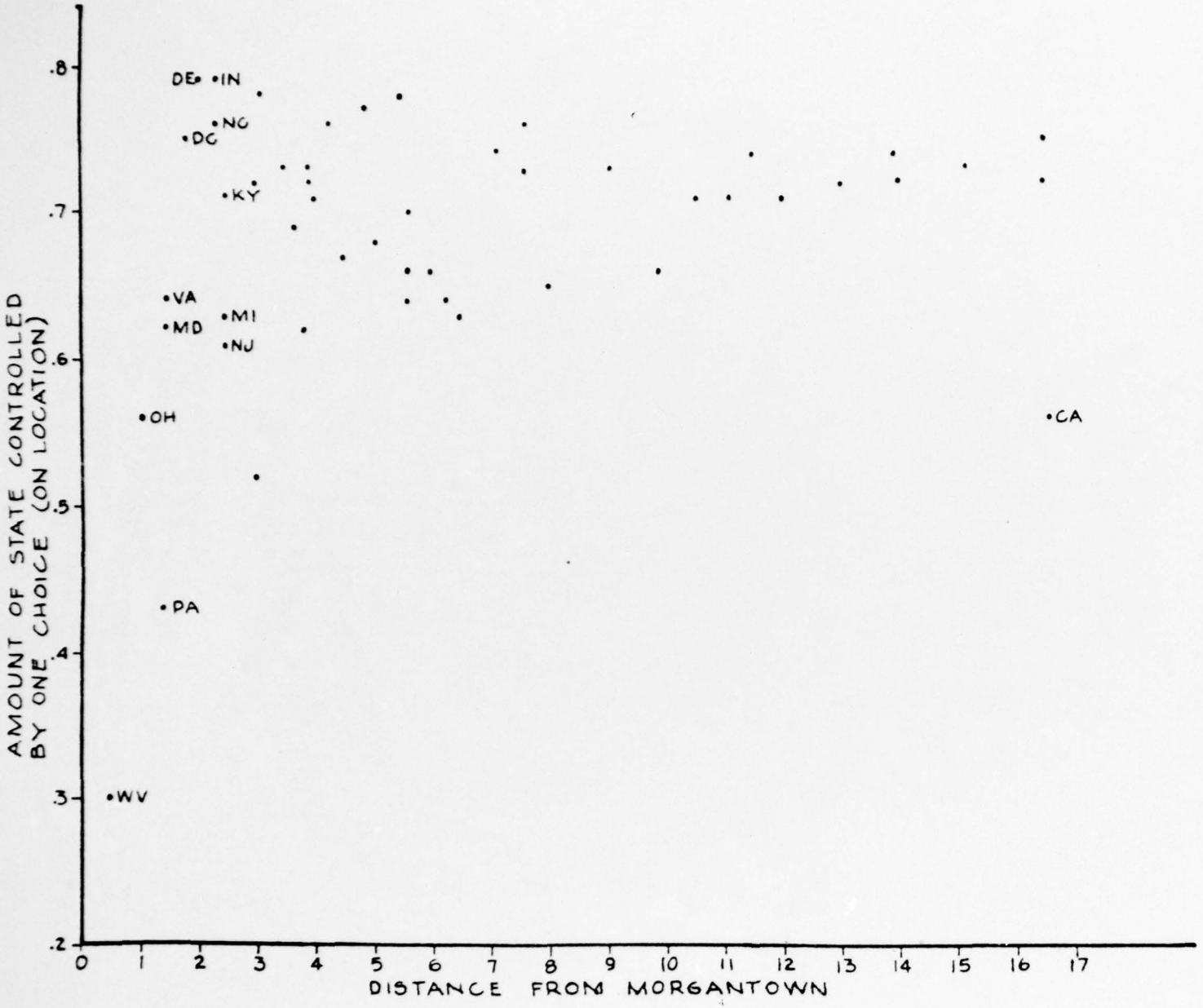


FIGURE 9. Fractional amount of each state controlled by the optional choice for each state, as a function of distance from Morgantown (the units are cm. on a map in a telephone directory).

FIGURE 10. Mean probability of an informant making a choice on the basis of occupation as a function of the occupational level of the trigger on the Duncan-Reiss scale. The straight line shows the best fit, accounting for nearly half the variance.

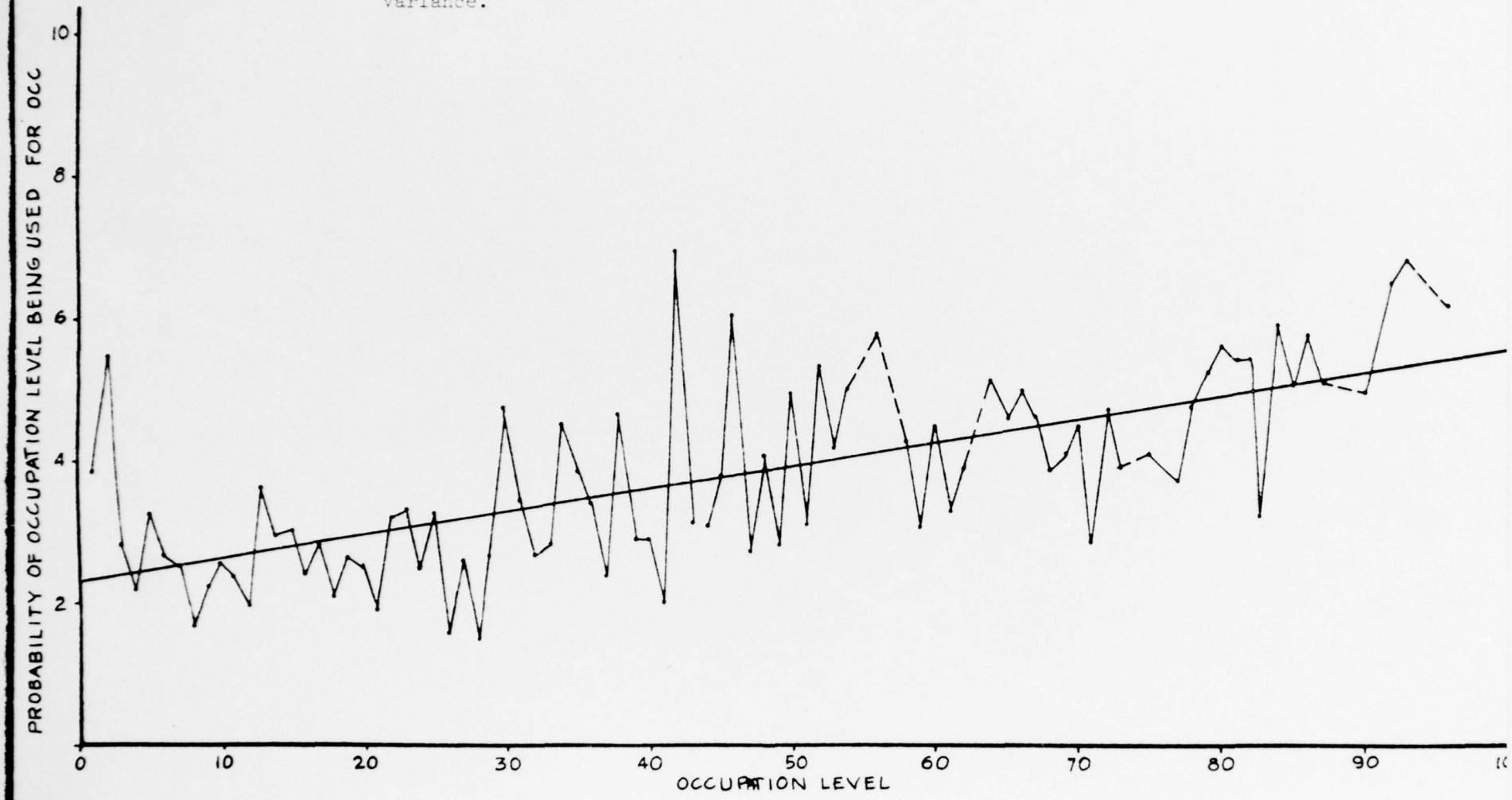


FIGURE 11. Number of choices estimated by informants compared to the number of choices they made.